

# Using the lasso method for space-time short-term wind speed predictions

September 10, 2015

Daniel Ambach<sup>1</sup> and Carsten Croonenbroeck<sup>2</sup>

## Abstract

Accurate wind power forecasts depend on reliable wind speed forecasts. Numerical Weather Predictions (NWP) utilize huge amounts of computing time, but still have rather low spatial and temporal resolution. However, stochastic wind speed forecasts perform well in rather high temporal resolution settings. They consume comparably little computing resources and return reliable forecasts, if forecasting horizons are not too long. In the recent literature, spatial interdependence is increasingly taken into consideration. In this paper we propose a new and quite flexible multivariate model that accounts for neighbouring weather stations' information and as such, exploits spatial data at a high resolution. The model is applied to forecasting horizons of up to one day and is capable of handling a high resolution temporal structure. We use a periodic vector autoregressive model with seasonal lags to account for the interaction of the explanatory variables. Periodicity is considered and is modelled by cubic B-splines. Due to the model's flexibility, the number of explanatory variables becomes huge. Therefore, we utilize time-saving shrinkage methods like lasso and elastic net for estimation. Particularly, a relatively newly developed iteratively re-weighted lasso and elastic net is applied that also incorporates heteroscedasticity. We compare our model to several benchmarks. The out-of-sample forecasting results show that the exploitation of spatial information increases the forecasting accuracy tremendously, in comparison to models in use so far.

**Keywords:** Wind speed; Forecasting; Periodic vector autoregressive model; Periodic B-splines; Iteratively re-weighted lasso method

## Addresses:

<sup>1</sup> **Corresponding Author:** Daniel Ambach, European University Viadrina, Chair of Quantitative Methods and Statistics, Post Box 1786, 15207 Frankfurt (Oder), Germany, Tel. +49 (0)335 5534 2983, Fax +49 (0)335 5534 2233, ambach@europa-uni.de.

<sup>2</sup> Carsten Croonenbroeck, European University Viadrina, Chair of Economics and Economic Theory (Macroeconomics), Post Box 1786, 15207 Frankfurt (Oder), Germany, Tel. +49 (0)335 5534 2701, Fax +49 (0)335 5534 72701, croonenbroeck@europa-uni.de.

# 1 Introduction

The progressing energy turnaround in Europe has one designated goal: Reducing the dependence on fossil energy and thus, increasing the fraction of energy production that is obtained from renewable sources. Other renewables like solar power or geothermal power aside, wind power is the most successful one, in terms of installed capacity as well as in the aggregated amount of power production. Contrary to fossil power however, wind power production is erratic and non-deterministic. Thus, accurate predictions are crucial for efficient market clearing as well as network dispatching, as Soman et al. (2010) and Croonenbroeck and Dahl (2014) point out. Short- to medium-term wind power forecasting (up to one day) is a wide field of research. These predictions are usually carried out by stochastic approaches, as discussed by Wu and Hong (2007).

Wind power prediction depends basically on wind speed predictions. However, wind power forecasting is a task on its own, i.e. power production depends on the turbine type, among other things, as Burton et al. (2011) indicate. Wind speed prediction models have been under constant development and improvement during the recent decade, at least. Early work is done by Haslett and Raftery (1989). They propose a long-memory autoregressive moving average model for wind speed. Ewing et al. (2006) investigate heteroscedasticity of high frequency data (15-minute observation frequency). Thus, Koopman et al. (2007) and Taylor et al. (2009) apply an ARFIMA-GARCH model (autoregressive fractionally integrated moving average with generalized autoregressive conditional heteroscedasticity) with a seasonal component in the explanatory variables.

Instead of univariate processes, more recently, researchers exploit spatial information of the data, as weather stations are available at close proximity to each other in many cases. ? develop a spatial regime-switching model. Furthermore, Zhu et al. (2014) provide an extended regime-switching approach. Šaltyte Benth and Šaltyte (2011) model the spatial dependence of the daily wind speed by a Gaussian random field. Aguera-Perez et al. (2013) and Santos-Alamillos et al. (2014) investigate the spatial structure of the data as well.

In this paper we come up with a very general and flexible model that uses spatial wind speed, wind direction, temperature and air pressure data. The interaction of the data is modelled by a periodic vector autoregressive model (SVAR) with explanatory variables (referred to as “X”), which are, in our case, just several periodic regressors. Contrary to classical Fourier modeling for the periodicity, we use more flexible B-splines. The heteroscedastic variance is modelled by a threshold autoregressive conditional heteroscedasticity (TARCH) model, also with explanatory variables (which are again, in our case, periodic regressors). In the end, our model is an SVARX-TARCHX model with adaptive lag selection. Due to the huge parameter space, classical maximum likelihood (ML) estimation would consume a lot of computing time. Instead, we use the popular shrinkage methodology, applied by means of the elastic net (see ?) and lasso (least absolute selection and shrinkage operator) method (see Tibshirani, 1996). In the context of autoregressive and time series models, Ren and Zhang (2010), ? and ? apply

lasso type and elastic net methods. Evans et al. (2014) use the lasso and several other empirical models to enhance the forecasting performance of a wind farm. The lasso and the elastic net are efficient ways to perform model selection and model fitting in one step. These algorithms operate quite fast and do not require any distributional assumption. Recently, Ziel et al. (2015) proposed an iteratively re-weighted lasso method which incorporates heteroscedasticity within the variance part.

We apply our model to a data set that consists of seven weather stations in Germany. The stations record weather data at a frequency of ten minutes. Per station, we have three full sample years, i.e. roughly 158,000 observations. With this data set, we calculate forecasts from our model and from several benchmark models, i.e. the persistence model, an AR, a VAR model and an ARFIMA-APARCH model (ARFIMA with asymmetric power generalized autoregressive conditional heteroscedasticity), as applied by Ambach and Schmid (2014). The novel multivariate approach is calculated by two different estimation procedures, the lasso method and the elastic net. We compare all models according to their forecasting accuracy and give a proposal of the best modeling approach.

The article is structured as follows. In Section 2, we briefly describe the data set and some general properties of the data set. Furthermore, the univariate and the benchmark models are described. The novel multivariate approach is introduced. Section 3 describes the in-sample results of our new modeling approach. The out-of sample results are provided in Section 4. Finally, Section 5 concludes.

## 2 Wind speed data and model description

The spatial area of investigated wind speed data is shown in Figure 1. In this article we focus on the most central station Lindenberg [5]. We use Müncheberg [2] as a reference station and for spatial information, also use its neighbours. They are situated in Eastern Germany in a region of rural plains. This region is perfect for wind parks. The data is measured by the “Deutscher Wetterdienst” (DWD) and reaches from January 2009 to December 2011. For model fitting, a time frame of two and a half years is used and the remaining months (July 2011 to December 2011) are used for out-of-sample forecasts. The wind speed  $(W_{m,t})_{m \in \{1, \dots, M\}, t \in \{1, \dots, T\}}$  is measured in  $m/s$  in a 10-minute interval for a station  $m$  at time  $t$ . Figure 2 shows the wind speed data, the corresponding histogram and the autocorrelation function (ACF) for stations Lindenberg and Müncheberg. The observed wind speed possesses strong (conditional) volatility and the histogram of the wind speed shows a non-negative and positively skewed distribution. Moreover, we observe the presence of autocorrelation, which is related to the high frequency of the wind speed data. The autocorrelation function shows periodic characteristics. To investigate the periods, we calculate the smoothed periodogram, which is given in Figure 3. The peaks in this picture depict a periodic behavior for several frequencies, which correspond to diurnal  $s_1 = 1/0.000019 = 52,560$  and annual  $s_2 = 1/0.00694 = 144$  periods (note that at a data frequency

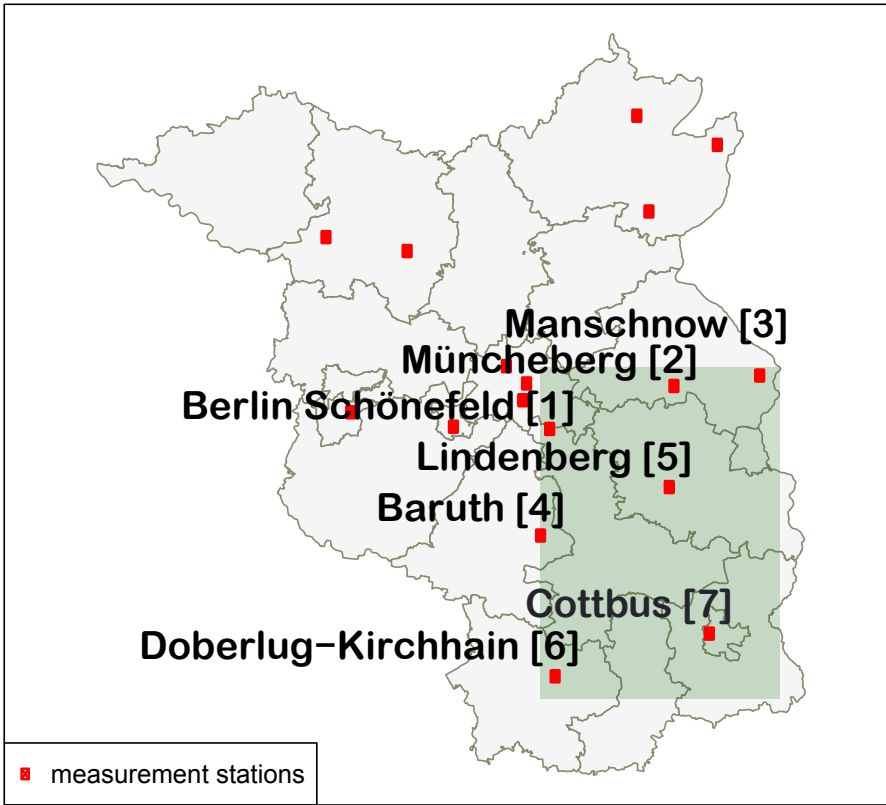


Figure 1. Meteorological measurement stations in Berlin and Brandenburg which provide 10min data.

of 10 minutes, there are 144 obs./day and 52,560 obs./year). We observe autocorrelation, a periodic behavior and heteroscedasticity, just as Ewing et al. (2006), Taylor et al. (2009) and ?. According to the fact that wind speed is a spatial phenomenon, it is reasonable to include the available and relevant information of neighbouring measurement stations. We will emphasize this idea with an example. If the wind is blowing from the north to the south and we observe this information at station Müncheberg, we are able to use this information for the station Lindenberg (cf. Figure 1). If the wind comes from the south, we can consider the observations from Cottbus. The investigated area of our data set contains seven different stations and Lindenberg is the midpoint.

Figure 4 presents the pairwise correlation of all variables of all stations for the entire in-sample time frame. Blue denotes positive correlation, red denotes negative correlation, darker colors indicate greater values. Crossed boxes represent insignificant correlation coefficients. It can be seen that aside from six exceptions, all correlation coefficients are significant, but some are close to zero. Müncheberg's air pressure is barely correlated to any of the other variables. Each station's temperature data is strongly positively correlated to each other station's temperature data. The same holds for the wind direction. The correlation between most groups of different types of variables is weak. Temperature and wind speed are weakly negatively correlated, wind speed and air pressure are somewhat stronger positively correlated. Eventually, we observe a huge positive correlation between the wind direction and the wind speed, which will be captured by our new multivariate model. Figure 5 depicts the wind rose for our target

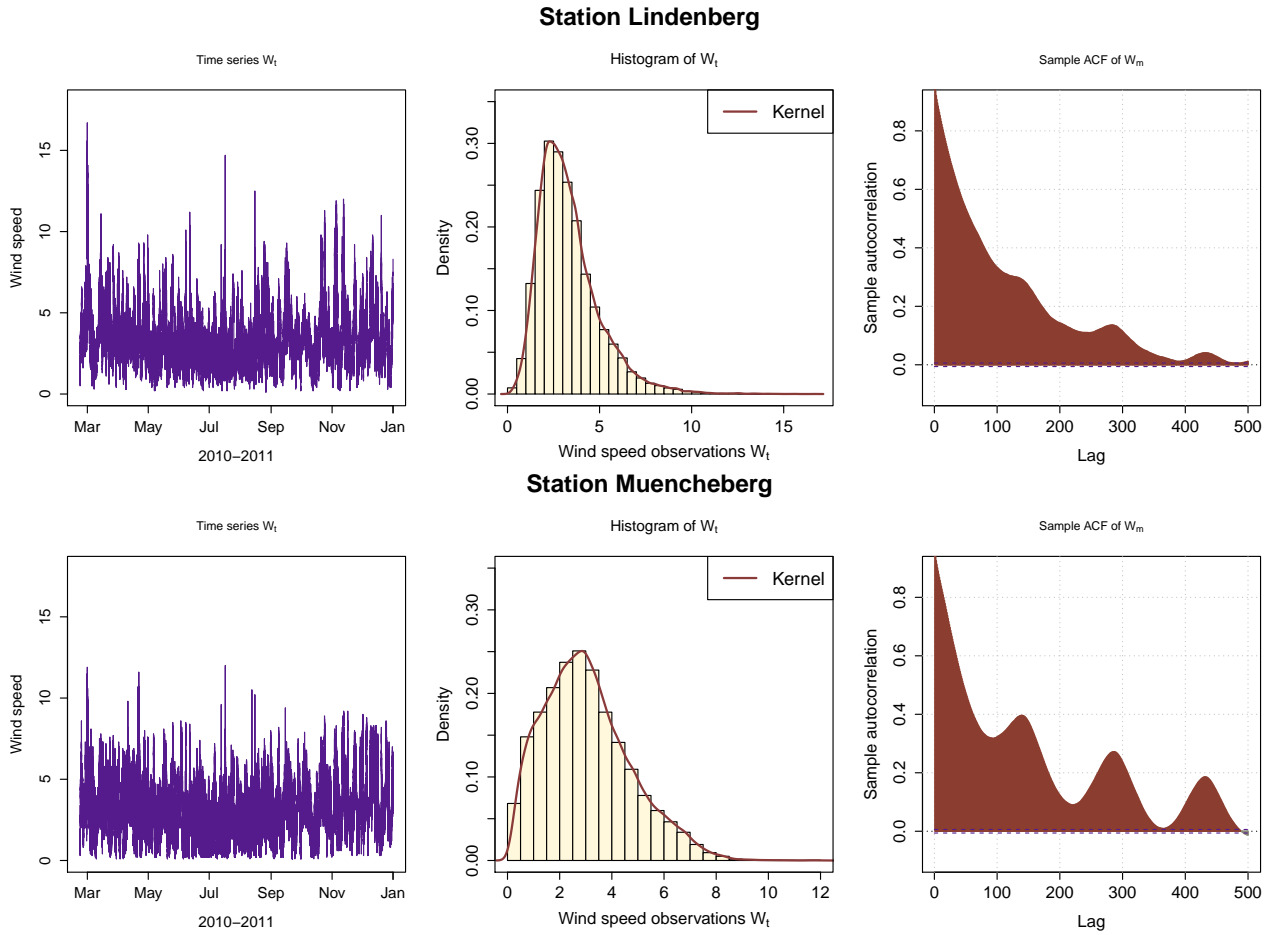


Figure 2. Wind speed data recorded at a 10-minute frequency, corresponding histogram and plot of the autocorrelation function.

station Lindenberg (left panel). Wind roses show the densities of wind speeds, dependent on the wind direction. As the wind speed densities are rather large for eastern, north-eastern and south-eastern directions, we conclude that Lindenberg is mostly affected by winds coming from directions at which Baruth, Doberlug-Kirchhain and possibly Berlin-Schönefeld are located. Accordingly, Figure 5 shows the wind rose for station Münchenberg on the right panel.

## 2.1 Benchmark models

Univariate wind speed models are based on a decomposition of the time series into a time-dependent intercept, a trend component and a seasonal part. A general model can be

$$W_t = \text{Intercept}_t + \text{Trend}_t + \text{Seasonal}_t + \epsilon_t, \quad (1)$$

where  $W_t$  represents the wind speed and  $\epsilon_t$  is the residual process with zero mean and time dependent variance  $\sigma_t$  for time  $t$ . One very general model of this type is the ARFIMA-APARCH approach with periodic regressors as proposed by Ambach and Schmid (2014), which nests various other models formerly proposed. As we contribute a novel multivariate model, we

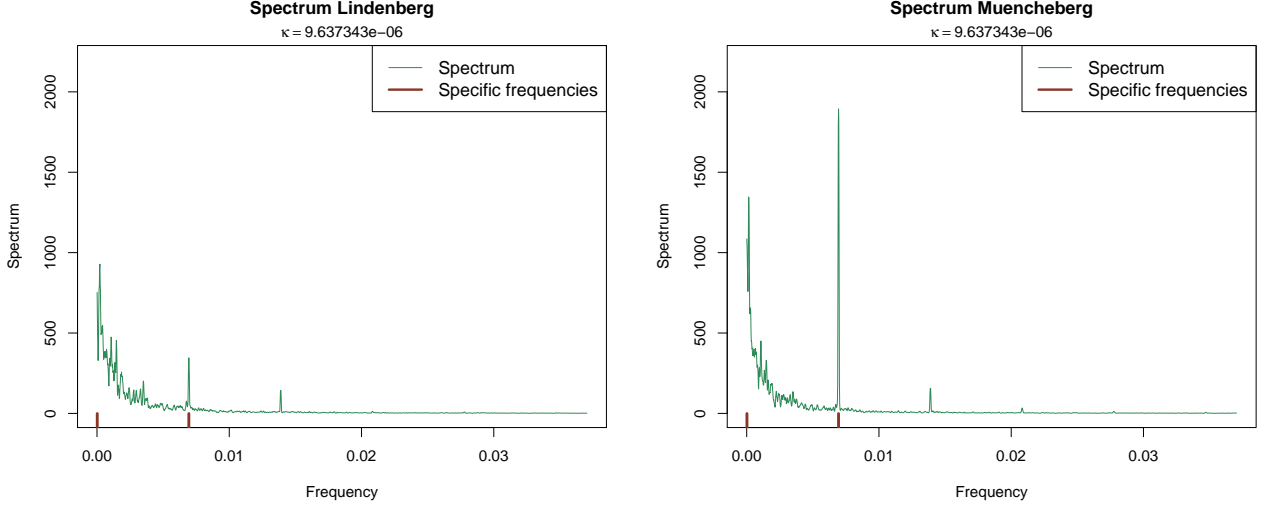


Figure 3. Estimation of the spectral density.

use the ARFIMA-APARCH merely as a competitive benchmark from the class of univariate models. As in the original motivation of this approach, we include periodic B-splines instead of Fourier series to capture periodicity/seasonality. The intercept and the seasonal components are modelled by

$$Intercept_t + Seasonal_t = \vartheta_{11} + \sum_{i_1=2}^{k_1} \sum_{i_2=2}^{k_2} \vartheta_{i_1, i_2} f_{i_1}^{s_1}(t) f_{i_2}^{s_2}(t). \quad (2)$$

where  $f_{i_1}^{s_1}(t)$  and  $f_{i_2}^{s_2}(t)$  are up to  $k_1$  and  $k_2$  time dependent periodic functions. The residual process  $\epsilon_t$  is stationary and follows a ARFIMA-APARCH. Thus, it is given by

$$\begin{aligned} \epsilon_t &\equiv \phi(B)(1-B)^d X_t = \theta(B) Z_t, \\ Z_t &= h_t \eta_t \quad \text{where} \quad \eta_t \stackrel{iid}{\sim} WN(0, \sigma^2) \\ h_t^\delta &= \alpha_0 + \sum_{l=1}^Q \alpha_l (|Z_{t-l}| - \gamma_l Z_{t-l})^\delta + \sum_{i=1}^P \beta_i h_{t-i}^\delta, \end{aligned} \quad (3)$$

where  $d \in (-0.5, 0.5)$  is the differencing parameter and  $B$  is the backward shift operator with  $B^u X_t = X_{t-u}$ . The ARMA(p,q) polynomials  $\phi$  and  $\theta$  are  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  and  $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  and have no common factors. The roots of  $\phi$  and  $\theta$  must lie outside the unit circle. ? point out that instead of the conditional variance, the conditional standard deviation with power  $\delta$  follows an asymmetric power ARCH (APARCH) process. The asymmetry parameter is  $\gamma_l \in [-1, 1]$  for  $l = 1, \dots, Q$ . ? provide the existence of a stationary solution for the APARCH process by  $\sum_{l=1}^Q \alpha_l E(|z| - \gamma_l z)^\delta + \sum_{i=1}^P \beta_i < 1$ .

The parameters of this ARFIMA-APARCH model are usually estimated by means of a quasi maximum likelihood (QML) estimation procedure, which can be slow, requires a strict distributional assumption and is prone to the usual problems of numeric solving, e.g. identification

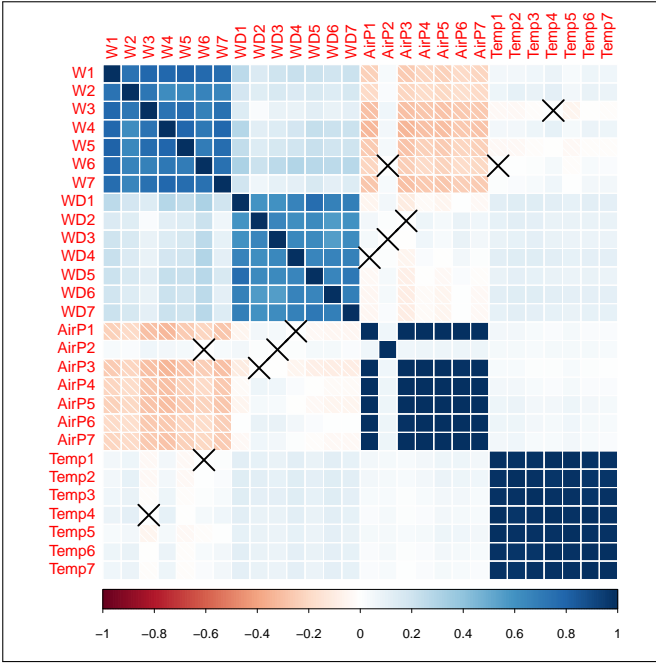


Figure 4. Plot of pairwise correlation for all dependent and independent variables and all stations, using plain sensor data.

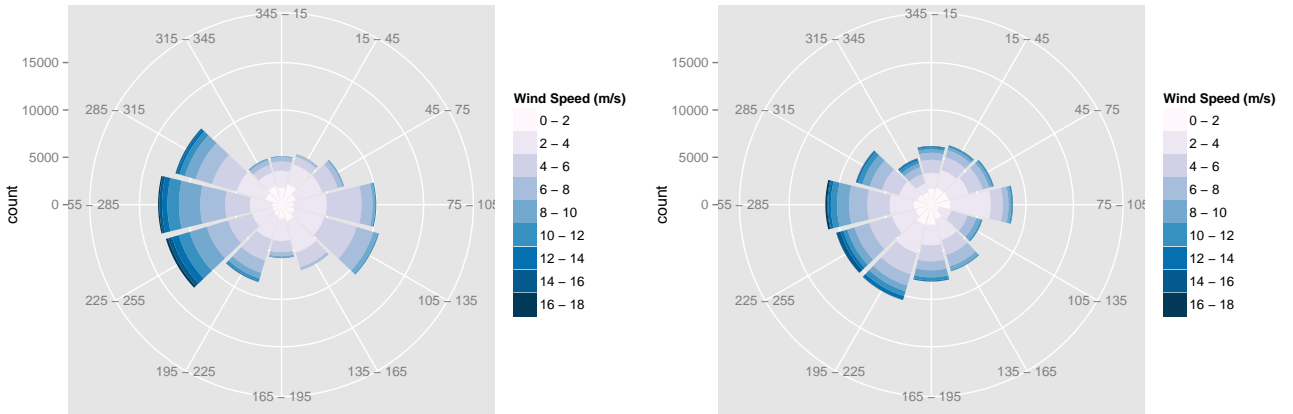


Figure 5. Wind rose, where the wind speed frequencies are plotted by wind direction for station Lindenberg (left) and station Müncheberg (right).

of the global maximum.

Ambach and Schmid (2014) describes seasonality modeling by means of periodic B-splines, which are becoming more and more accepted in the literature. Thapar et al. (2011), Bazilevs et al. (2012) and Le Guyader et al. (2014) show the power of this approach. The fundamental basics of the spline functions are defined by ? and ?. In the application, we define our B-Splines by a set of equidistant knots such that each function oscillates once per day. Also, we use a second setup of functions to capture annual periodicity. All B-Splines are twice continuously differentiable. Figure 6 depicts our diurnal functions.

Aside from the nested ARFIMA-APARCH model for  $\epsilon_t$ , we consider a simple  $VAR(p)$  (vector autoregressive) and an  $AR(p)$  process. The vector contains wind speed, but also several other variables, which are described in Section 2.2. Finally, we investigate the usual persistence or

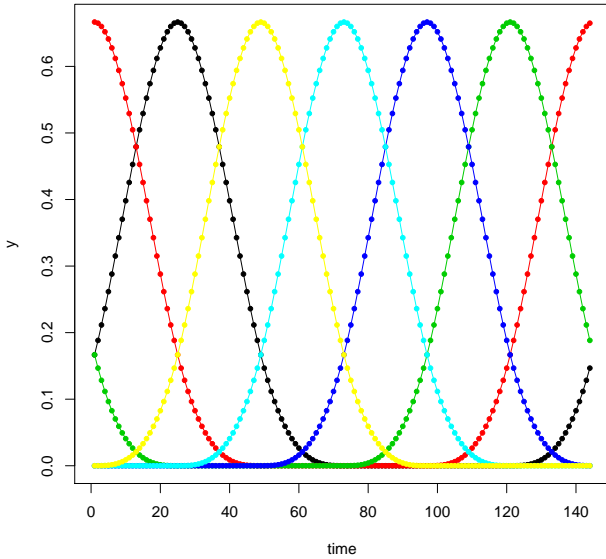


Figure 6. Diurnal cubic B-splines.

“naïve” benchmark (see ?). The naïve predictor uses the current value as a forecast for future values  $\widehat{W}_{t+k} = W_t$ .

## 2.2 Novel multivariate wind speed prediction model

In contrast to the established univariate modeling by, e.g., the above-mentioned ARFIMA-APARCH process, we introduce a multivariate model. Basically, our model is a VAR, but the mean is enriched by seasonal functions and the conditional variance of the residual process is modelled by a threshold autoregressive process (TARCH). The mean part as well as the variance part may contain external periodic regressors (X), so the model can be described as an SVARX-TARCHX approach.

We decompose our wind direction information into an east-west component, given by  $\sin(az_{m,t})$ , and a north-south component, given by  $\cos(az_{m,t})$ , where  $az_{m,t}$  is the azimuth, as measured at station  $m$  for time  $t$ . The SVARX part captures wind speed information, but also, air pressure information  $AP_{m,t}$  and temperature  $C_{m,t}$ , measured in degrees Celsius are incorporated. For stations  $m = 1, \dots, M$  and time point  $t = 1, \dots, T$ , the dependent vector consists of observations for 5 variables, i.e.  $\mathbf{Y}_t \in \mathbb{R}^{5 \cdot M}$ , which is

$$\mathbf{Y}_t = \begin{pmatrix} W_{1,t}, \dots, W_{M,t}, \sin(az_{1,t}), \dots, \sin(az_{M,t}), \\ \cos(az_{1,t}), \dots, \cos(az_{M,t}), AP_{1,t}, \dots, AP_{M,t}, C_{1,t}, \dots, C_{M,t} \end{pmatrix}'. \quad (4)$$

The model is accounting for an overall relationship between speed and direction variables across all sites. The multivariate dependent variables in  $\mathbf{Y}_t$  are described by a vector autoregressive model for the mean part given by

$$\mathbf{Y}_t = \boldsymbol{\vartheta}_t + \sum_{j=1}^J \boldsymbol{\phi}_{t,j} \mathbf{Y}_{t-j} + \boldsymbol{\epsilon}_t, \quad (5)$$

$$\boldsymbol{\vartheta}_t = \boldsymbol{\vartheta}_0 + \sum_{i_1=2}^{k_1} \boldsymbol{\vartheta}_{i_1,1} f_{i_1}^{s_1}(t) + \sum_{i_2=2}^{k_2} \boldsymbol{\vartheta}_{1,i_2} f_{i_2}^{s_2}(t) + \sum_{i_1=2}^{k_1} \sum_{i_2=2}^{k_2} \boldsymbol{\vartheta}_{i_1,i_2} f_{i_1}^{s_1}(t) f_{i_2}^{s_2}(t), \quad (6)$$

$$\boldsymbol{\phi}_{t,j} = \boldsymbol{\phi}_{0,j} + \sum_{i_1=2}^{k_1} \boldsymbol{\phi}_{i_1,1,j} f_{i_1}^{s_1}(t) + \sum_{i_2=2}^{k_2} \boldsymbol{\phi}_{1,i_2,j} f_{i_2}^{s_2}(t) + \sum_{i_1=2}^{k_1} \sum_{i_2=2}^{k_2} \boldsymbol{\phi}_{i_1,i_2,j} f_{i_1}^{s_1}(t) f_{i_2}^{s_2}(t), \quad (7)$$

$$\boldsymbol{\epsilon}_t = \boldsymbol{\sigma}_t \boldsymbol{\eta}_t, \quad (8)$$

where  $\{\boldsymbol{\eta}_t\}$  is i.i.d.,  $E(\boldsymbol{\eta}_t) = \mathbf{0}$  and  $\text{Var}(\boldsymbol{\eta}_t) = \mathbf{1}$ . Moreover,  $\boldsymbol{\vartheta}_0$  is an  $(M \cdot 5) \times 1$  intercept vector and  $\boldsymbol{\vartheta}_{i_1,1}$ ,  $\boldsymbol{\vartheta}_{1,i_2}$  and  $\boldsymbol{\vartheta}_{i_1,i_2}$  are  $(M \cdot 5) \times 1$  periodic coefficient vectors.  $\boldsymbol{\phi}_{0,j}$ ,  $\boldsymbol{\phi}_{i_1,1,j}$ ,  $\boldsymbol{\phi}_{1,i_2,j}$  and  $\boldsymbol{\phi}_{i_1,i_2,j}$  are  $(M \cdot 5) \times (M \cdot 5)$  parameter matrices of autoregressive and periodic autoregressive parameters for lag  $j \in \mathbb{N}$ . The periodic B-spline functions  $f_{i_1}^{s_1}(t)$  and  $f_{i_2}^{s_2}(t)$  are similar to Section 2.1. Finally,  $\boldsymbol{\sigma}_t$  and  $\boldsymbol{\eta}_t$  are  $(M \cdot 5) \times 1$  vectors which follow a TARCH process (see Glosten et al., 1993) given by

$$\boldsymbol{\sigma}_t = \boldsymbol{\alpha}_0 + \sum_{i_1=2}^{k_1} \boldsymbol{\alpha}_{i_1} f_{i_1}^{s_1}(t) + \sum_{h=1}^P \boldsymbol{\zeta}_{t,h} \mathbf{I}_t^+ \boldsymbol{\epsilon}_{t-h} + \sum_{l=1}^Q \boldsymbol{\psi}_{t,l} \mathbf{I}_t^- \boldsymbol{\epsilon}_{t-l}, \quad (9)$$

$$\boldsymbol{\zeta}_{t,h} = \boldsymbol{\zeta}_{0,h} + \sum_{i_1=2}^{k_1} \boldsymbol{\zeta}_{i_1,h} f_{i_1}^{s_1}(t), \quad (10)$$

$$\boldsymbol{\psi}_{t,l} = \boldsymbol{\psi}_{0,l} + \sum_{i_1=2}^{k_1} \boldsymbol{\psi}_{i_1,l} f_{i_1}^{s_1}(t), \quad (11)$$

where  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\alpha}_{i_1}$  are  $(M \cdot 5) \times 1$  parameter vectors. Furthermore,  $\boldsymbol{\zeta}_{0,h}$ ,  $\boldsymbol{\psi}_{0,l}$ ,  $\boldsymbol{\zeta}_{i_1,h}$  and  $\boldsymbol{\psi}_{i_1,l}$  are autoregressive and periodic autoregressive parameter matrices  $((M \cdot 5) \times (M \cdot 5))$  within the variance. Hence, the  $M \cdot 5 \times 1$  vectors of indicator functions  $\mathbf{I}_t^+$  and  $\mathbf{I}_t^-$  are given by

$$\mathbf{I}_{\mathcal{N},m,t-1}^+ = \begin{cases} 1, & \epsilon_{\mathcal{N},m,t-1} > 0 \\ 0, & \epsilon_{\mathcal{N},m,t-1} \leq 0 \end{cases}, \quad \mathbf{I}_{\mathcal{N},m,t-1}^- = \begin{cases} 1, & \epsilon_{\mathcal{N},m,t-1} \leq 0 \\ 0, & \epsilon_{\mathcal{N},m,t-1} > 0 \end{cases}, \quad (12)$$

where  $\mathcal{N} \in \{1, \dots, 5\}$  represents the  $\mathcal{N}$ th regressor. These vectors provide one way to model a TARCH process for a station  $m$ . Therefore, we double the parameter space, but are able to differentiate between negative and positive shocks. After all, the number of parameters amounts to

$$2 \cdot (J + 2((k_1 - 1) + (k_2 - 1) + (k_1 k_2 - 1) + 1) + (k_1 - 1) + P + Q + 1 + 2k_1),$$

which may be much larger than the number of observations. Thus, regularization and shrinkage are necessary.

? use an iteratively re-weighted least squares approach to develop accurate load forecasts. More recently, Ziel et al. (2015) implement an iteratively re-weighted lasso method to predict the electricity price. We take the lasso and the elastic net as estimation methods into consideration, as they are designed to reduce the parameter space in the estimation stage. Ren and Zhang (2010) take the lasso method in the context of vector autoregressive models. The weighted lasso estimation of the parameter vector  $\boldsymbol{\theta}_{\mathcal{N},m}$  for both the SVARX model (5) and the TARCHX model (9) is given by

$$\hat{\boldsymbol{\theta}}_{\mathcal{N},m} = \arg \min_{\boldsymbol{\theta}_{\mathcal{N},m} \in \mathbb{R}^{p_j}} \sum_{t=1}^T (Y_{\mathcal{N},m,t} - \omega_t \mathbf{X}_{\mathcal{N},m,t} \boldsymbol{\theta}_{\mathcal{N},m,t})^2 + \lambda_{\mathcal{N},m,T} \sum_{j=1}^{p_j} |\boldsymbol{\theta}_j|, \quad (13)$$

where  $\omega_t$  is the weight vector  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_T)$ ,  $p_j$  being the number of elements of the parameter vector  $\boldsymbol{\theta}_{\mathcal{N},m}$  and the tuning parameter is  $\lambda_{\mathcal{N},m,T} \geq 0$ . The second estimation method is the elastic net (see ?) which is given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\mathcal{N},m} = \arg \min_{\boldsymbol{\theta}_{\mathcal{N},m} \in \mathbb{R}^{p_j}} & \sum_{t=1}^T (Y_{\mathcal{N},m,t} - \omega_t \mathbf{X}_{\mathcal{N},m,t} \boldsymbol{\theta}_{\mathcal{N},m,t})^2 + \lambda_{\mathcal{N},m,T} \alpha \sum_{j=1}^{p_j} |\boldsymbol{\theta}_j| \\ & + \frac{1}{2} \lambda_{\mathcal{N},m,T} (1 - \alpha) \sum_{j=1}^{p_j} \boldsymbol{\theta}_j^2, \end{aligned} \quad (14)$$

where  $\alpha \in [0, 1]$  provides an additional tuning parameter. In the case of  $\alpha = 1$ , we obtain the lasso case and if  $\alpha = 0$ , we obtain the ridge regression. Finally, if  $\alpha \in (0, 1)$ , the penalties represent the elastic net.

The estimation is done in two-steps. We start with the estimation of the mean model and afterwards, we re-weight the mean model using  $\hat{\boldsymbol{\sigma}}_t$ . Figure 7 provides the estimation scheme for the iteratively re-weighted lasso method (see Efron et al., 2004, Ziel et al., 2015). During the algorithm, we use the Akaike information criterion (AIC) for the model selection. The procedure is repeated until some convergence is achieved. If  $\mathcal{K} = 1$ , we obtain the homoscedastic estimates without re-weighting the lasso. The advantage of this approach is the fast computing time, and further extensions are still possible. Clearly, we have to choose the specific lags of the AR and the TARCH part, which we discuss in the next section.

From the modeling perspective, we expect that the iteratively re-weighted lasso or elastic net method should be superior compared to the ML/QML approaches, because we are able to handle a huge parameter space with a variable selection and optimization algorithm. Moreover, we do not need a specific distributional assumption.

### 3 Model fitting results

The ARFIMA-APARCH process presented in Equations (2) and (3) is our extensive benchmark model. Here, this model is estimated by a QML approach under normally distributed residuals.

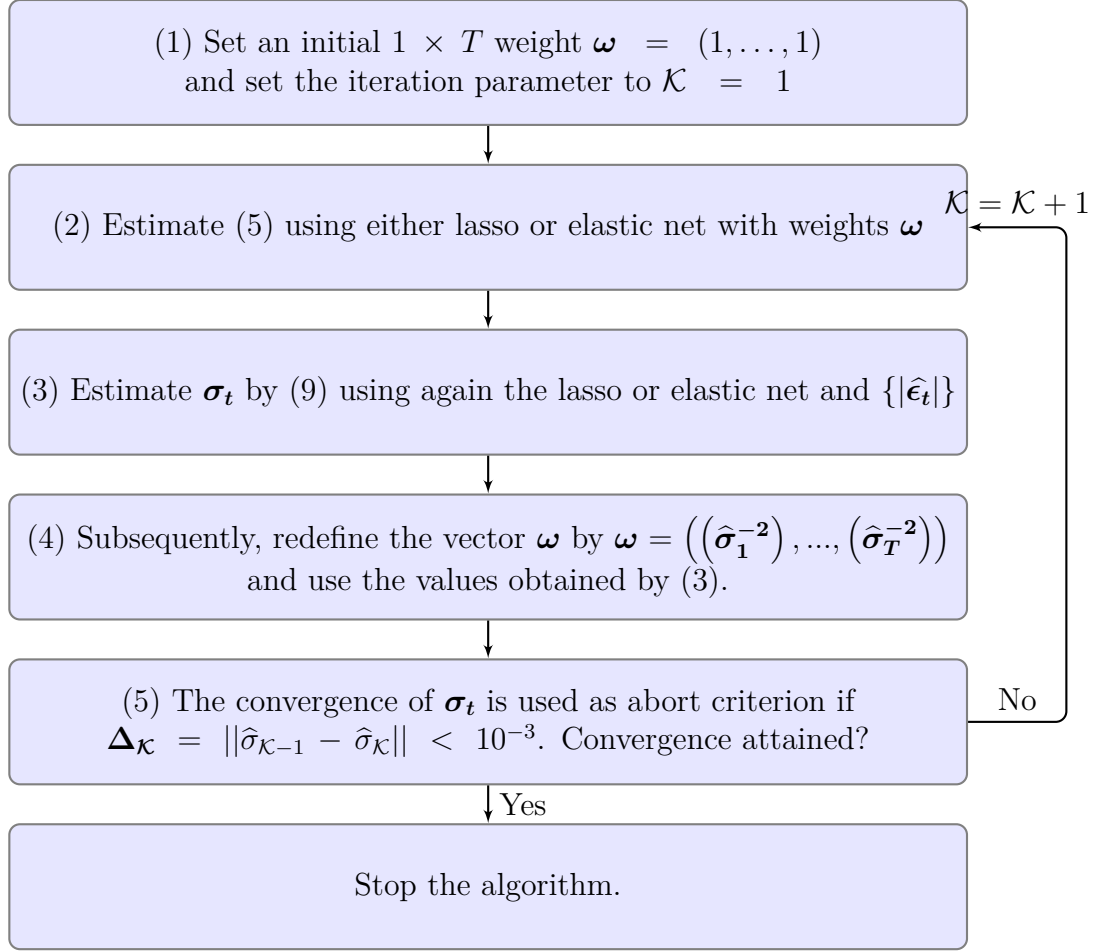


Figure 7. Estimation scheme.

According to the complex ARFIMA-APARCH, we determine a sparse parametrization. The in-sample results are comparable to previous findings by Ambach and Schmid (2014) and Taylor et al. (2009).

The novel wind speed approach is described in Equations (5) and (9). The estimation of the model is done by an iteratively re-weighted method. We distinguish between two estimation methods, the lasso method (13) and the elastic net (14). An advantage of these methods is that in contrast to the QML estimation of the ARFIMA-APARCH benchmark model, we do not need a distributional assumption for the SVARX-TARCHX model.

The wind speed time series shows a huge presence of autocorrelation with a strong diurnal structure. Therefore, we choose to take an autocorrelation structure of about two days, which is  $J = 288 + 1$ . The TARCH conditional variance structure of our model includes a periodicity of two days, i.e.  $Q = P = 288 + 1$  lags. The covariance structure between each station and each response variable is also modelled by the vector autoregressive lags within mean and variance part, but some interactions may be set to zero during the lasso/elastic net iterations.

Figure 8 depicts the autocorrelation functions of station Müncheberg and Lindenberg for the standardized residuals of the wind speed. Each  $2 \times 2$  block represents the respective angle on the diagonals: The main diagonal (top-left and bottom-right) represents the ACF of stations 2 and 5. The off-diagonal (top-right and bottom-left) represents cross-correlations: Top-right represents the ACF of station 2 on station 5, bottom-left represents ACF of station 5 on station 2. Quite few of them (in all cases, fewer than 5 %) are significant.

Figure 9 depicts the autocorrelation functions of station Müncheberg and Lindenberg for the

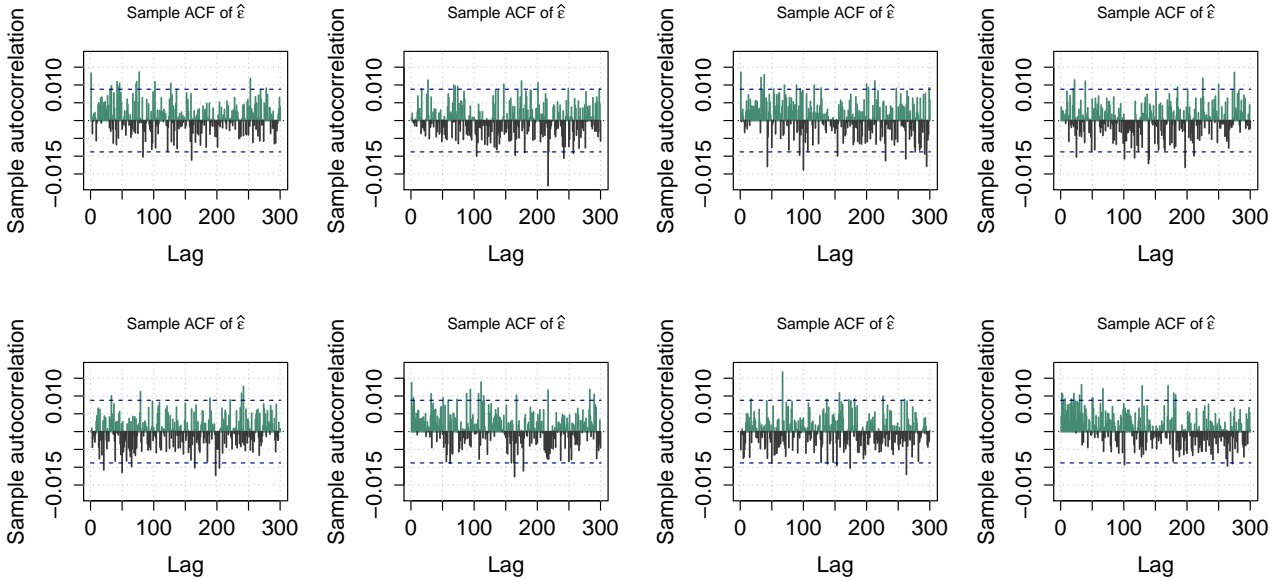


Figure 8. Autocorrelation function (ACF) of  $\{\hat{\epsilon}_t\}$  for the new model with lasso method (first and second column) and ACF of the elastic net (third and forth column).

absolute standardized residuals of the wind speed series  $\{|\hat{\epsilon}_t|\}$  and both estimation methods. The ACFs in Figures 8 and 9 show only a minor presence of remaining autocorrelation, as there are very few single spikes outside the confidence bands.

Additionally to ACF plots, we calculate the Ljung-Box test for  $\{\hat{\epsilon}_t\}$  and  $\{|\hat{\epsilon}_t|\}$ . Applying a

level of significance 5%, we cannot reject the null hypothesis of independence. After all, the autocorrelation analysis suggests an excellent model fit, especially due to the fact that almost no periodic structure remains in the residuals. Therefore, we expect proper forecasting results.

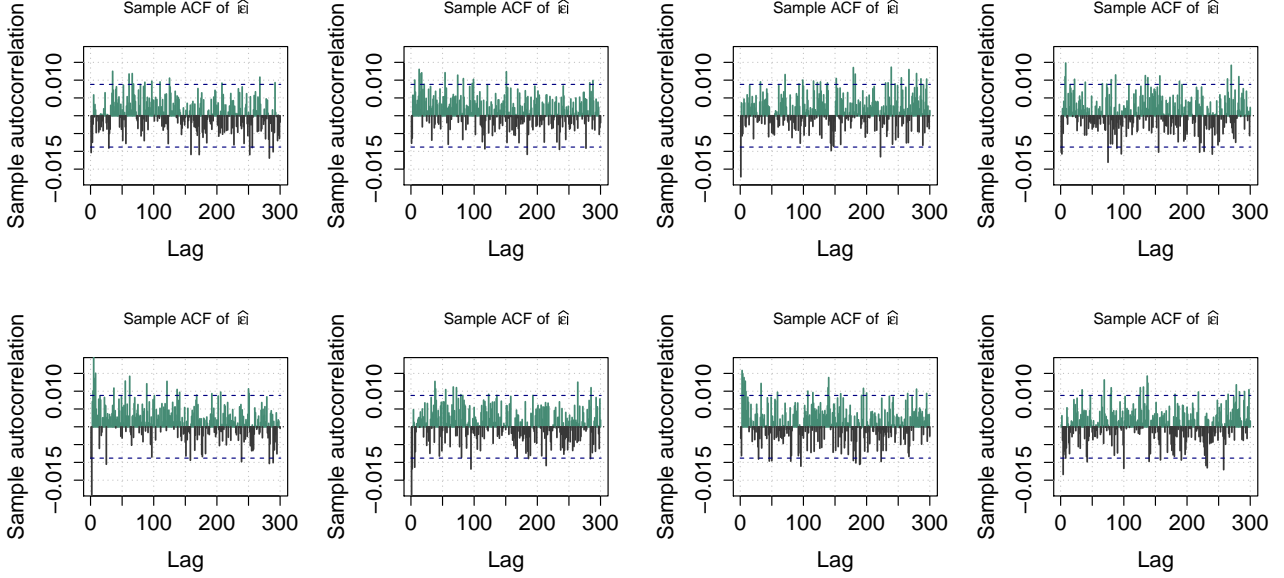


Figure 9. Autocorrelation function of  $\{|\hat{\epsilon}_t|\}$  (ACF) for the new model with lasso method (first and second column) and ACF of the elastic net (third and fourth column).

## 4 Out-of-sample forecasting results

In this section, we evaluate our model and the benchmark approaches according to their prediction performance. Common criteria are the root mean square error (RMSE), the mean absolute error (MAE) and the probability integral transform (PIT) histogram. PIT histograms allow for an insight into calibration and sharpness of forecasts, as ? point out. The forecasts are performed for a time frame from July 2011 to December 2011 for out-of-sample forecasts. We select  $N = 1000$  points in time  $(\tau^{(i)}, i = 1, \dots, N)$  in the out-of-sample period at random. Forecasts are calculated at horizons of up to a maximum of one day (i.e. 24 hours = 144 steps), as 24 hours conclude the short- to medium-term forecasting horizon of wind speed, as, e.g., ? and ? point out. RMSE and MAE are calculated by

$$RMSE_o = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( Y_{\mathcal{N}, m, \tau^{(i)} + o} - \hat{Y}_{\mathcal{N}, m, \tau^{(i)} + o} \right)^2}, \quad (15)$$

$$MAE_o = \frac{1}{N} \sum_{i=1}^N \left| Y_{\mathcal{N}, m, \tau^{(i)} + o} - \hat{Y}_{\mathcal{N}, m, \tau^{(i)} + o} \right|, \quad (16)$$

where  $\hat{Y}_{\mathcal{N},m,\tau^{(i)}+o}$  is the  $o$ -step forecast of wind speed and  $Y_{\mathcal{N},m,\tau^{(i)}+o}$  is the actual observation. Figures 10 and 11 present the out-of-sample aggregated forecasting error results. In all cases, our novel model is able to outperform the benchmark models. Mostly, even the competitive ARFIMA and the extensive VAR(p) model are outperformed substantially. Additionally, we observe that the naïve model is clearly outperformed in each case. For few shorter forecasting steps, the highly persistent VAR(p) model returns weakly lower errors. Indeed, these results are only obtained for the station Lindenberg. As forecasting horizons increase beyond one hour, our novel model is the overall winner. Finally, we are able to conclude that the iteratively re-weighted lasso and elastic net approach outperform the benchmark methods. Looking at MAE, the elastic net outperforms lasso. RMSE however leaves no distinct result for lasso and elastic net.

Figures 12 and 13 show the PIT histograms for shorter-term forecasting horizons of one and

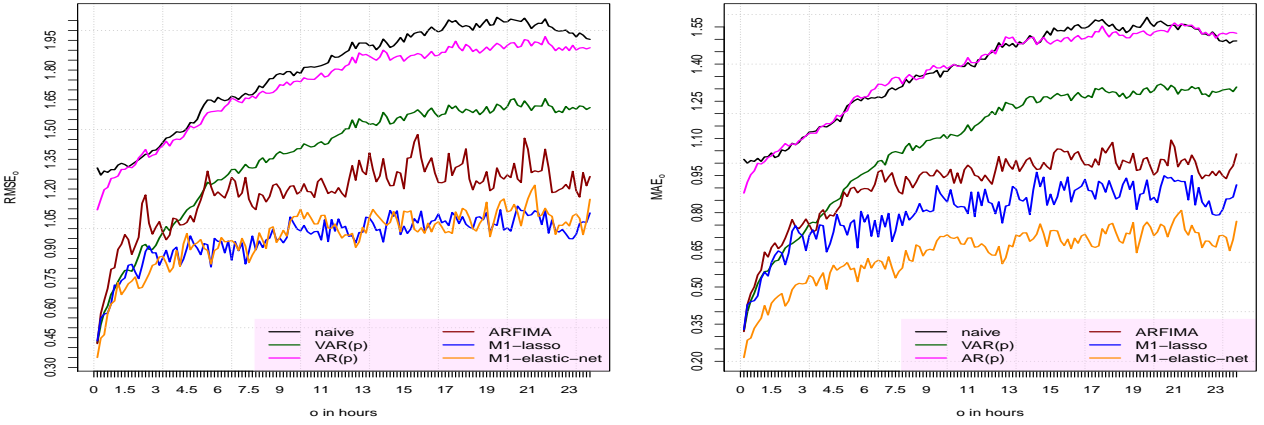


Figure 10. RMSE (first column) and MAE (second column) by forecasting horizon for station Müncheberg for all models.

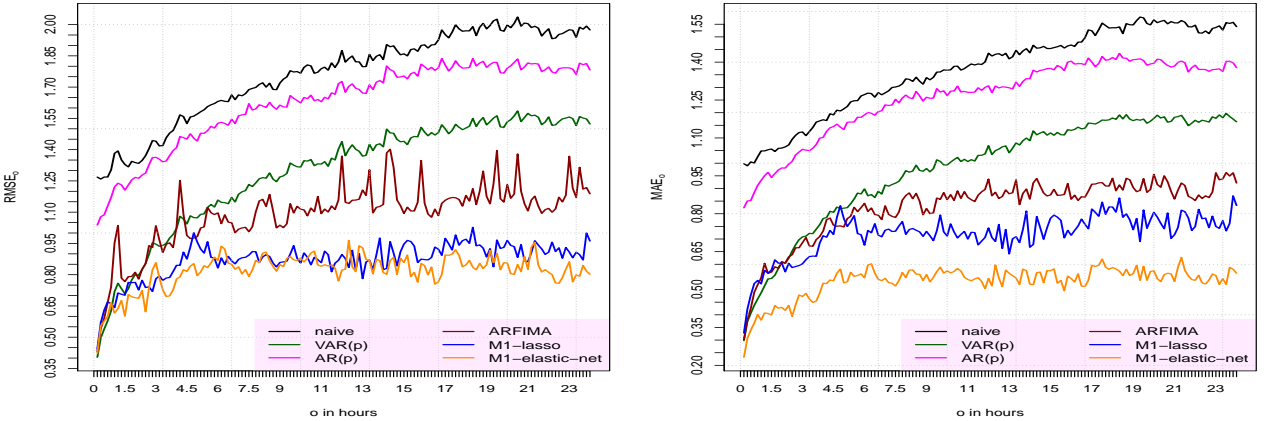


Figure 11. RMSE (first column) and MAE (second column) by forecasting horizon for station Lindenberg for all models.

six hours and the iteratively re-weighted lasso and elastic net forecasting technique. As to be expected, we observe that the calibration and sharpness of our forecasts is stable as long as forecasting horizons remain relatively short. However, we observe declining calibration and sharpness for forecasting horizons beyond a scope of six hours. So for medium-term forecasts,

we observe a PIT histogram which shows a certain degree of over-dispersion (omitted here to conserve space, available upon request).

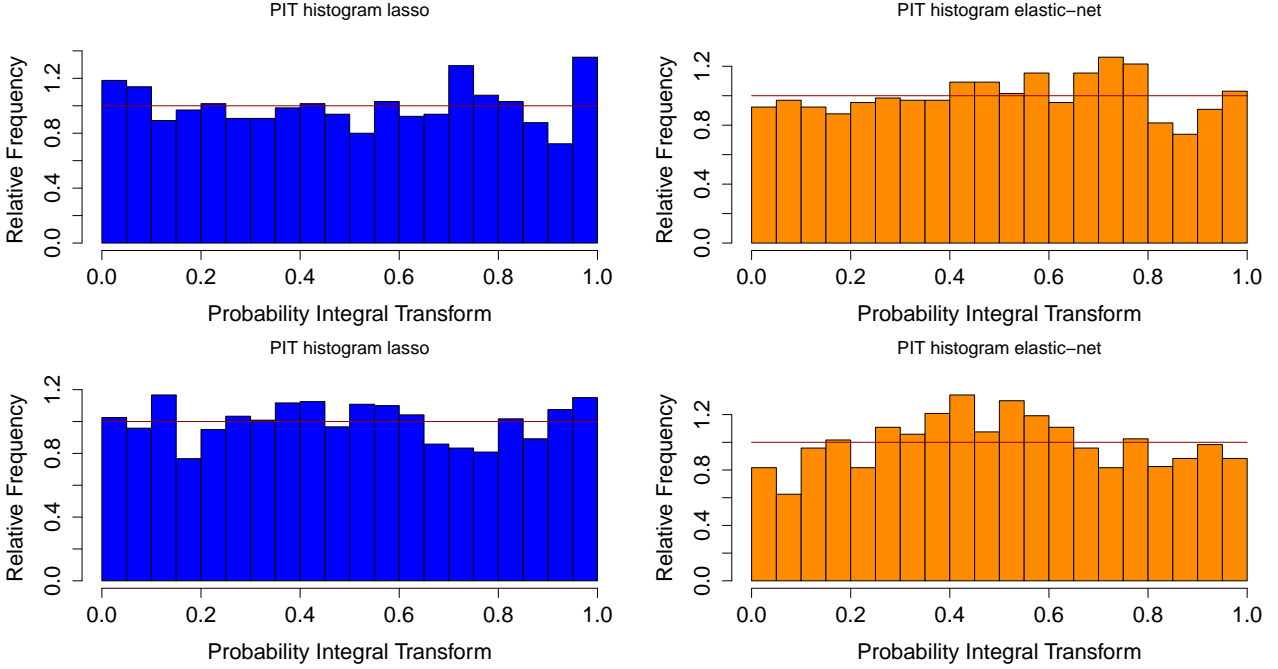


Figure 12. Probability integral transform (PIT) histogram for lasso (first column) and elastic-net (second column) for the forecasting horizon of one hour for station Müncheberg (first row) and Lindenberg (second row).

## 5 Conclusion

This article presents a new model for wind speed forecasting. The introduced periodic SVARX-TARCHX model is used to predict the wind speed, wind direction, air pressure and temperature. The model does so by means of non-trigonometric periodic B-spline functions. Moreover, we exploit the spatial distribution of the measurement stations and take conditional heteroscedasticity into account.

As the parameters of the sophisticated ARFIMA-APARCH benchmark are estimated by using numerical (quasi) maximum likelihood estimation, it usually takes more than one hour of computing time to find one out-of-sample forecast. In contrast to that, our SVARX-TARCHX model can be identified by modern shrinkage methods like lasso or elastic net, which takes only about six to eight minutes per forecast. Due to our iterative re-weighting scheme, the algorithm is still able to capture the heteroscedastic nature of the observed data.

Overall, our model proposition is able to outperform the naive benchmark as well as the VAR. Even the more advanced ARFIMA-APARCH model is outperformed by a significant degree. Modeling periodicity by B-splines instead of Fourier series brings additional flexibility and numerical performance. Results show that the new model captures periodicity quite well. The new model provides a flexible framework with a lot of adjustment options, which makes it a universal tool for many kinds of settings in wind speed forecasting research.

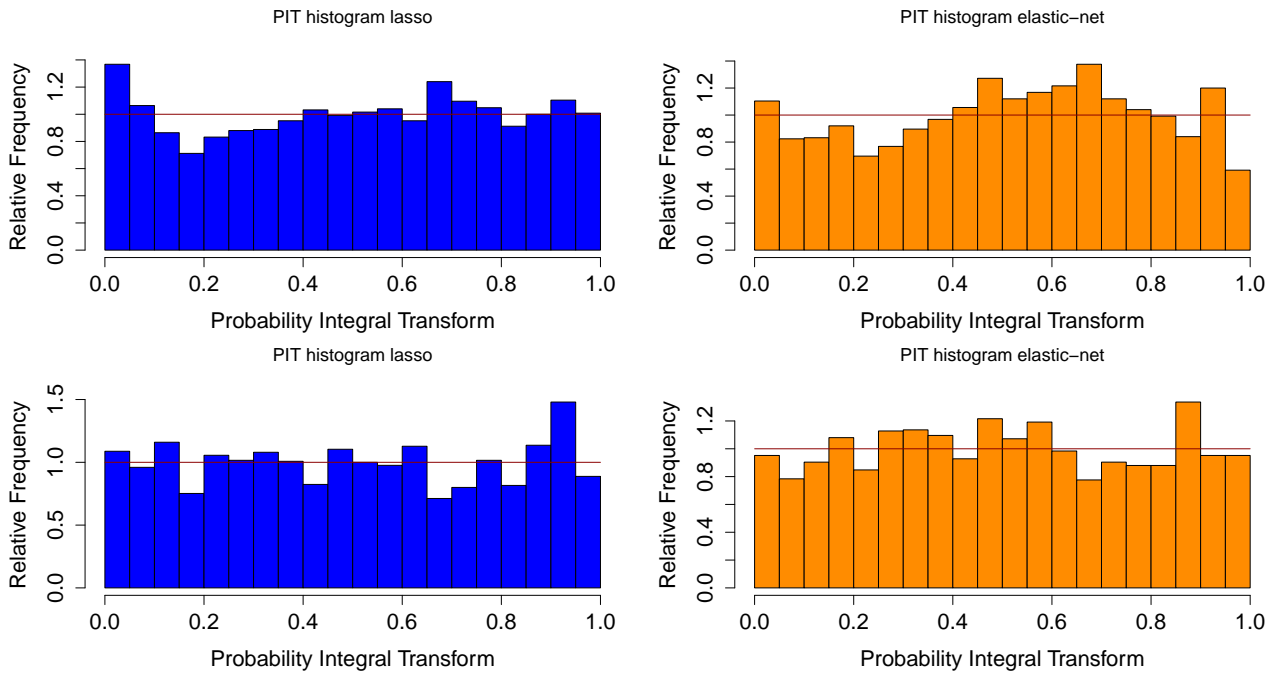


Figure 13. Probability integral transform (PIT) histogram for lasso (first column) and elastic-net (second column) for the forecasting horizon of four hours for station Müncheberg (first row) and Lindenberg (second row).

## References

- Aguera-Perez, A., Palomares-Salas, J.C., Gonzalez de la Rosa, J.J., and Moreno-Munoz, A. (2013), Spatial persistence in wind analysis, *Journal of Wind Engineering and Industrial Aerodynamics*, 119, pp. 48–52.
- Ambach, D. and Schmid, W. (2014), Periodic and Long Range Dependent Models for High Frequency Wind Speed Data, European University Viadrina Frankfurt (Oder), Department of Business Administration and Economics, Discussion Paper No. 358.
- Bazilevs, Y., Hsu, M., and Scott, M. (2012), Isogeometric Fluid-Structure Interaction Analysis with Emphasis on Non-Matching Discretizations, and with Application to Wind Turbines, *Computer Methods in Applied Mechanics and Engineering*, 249-252, pp. 28–41.
- Benth, J.Š. and Benth, F.E. (2010), Analysis and modelling of wind speed in New York, *Journal of Applied Statistics*, 37(6), pp. 893–909.
- Burton, T., Jenkins, N., Sharpe, D., and Bossanyi, E. (2011), *Wind Energy Handbook*, John Wiley & Sons, West Sussex, UK.
- Croonenbroeck, C. and Dahl, C.M. (2014), Accurate medium-term wind power forecasting in a censored classification framework, *Energy*, 73, pp. 221–232.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), Least angle regression, *The Annals of statistics*, 32(2), pp. 407–499.

- Evans, S.C., Zhang, Z., Iyengar, S., Chen, J., Hilton, J., Gregg, P., Eldridge, D., Jonkhof, M., McCulloch, C., and Shokoohi-Yekta, M. (2014), Towards wind farm performance optimization through empirical models, *Aerospace Conference, 2014 IEEE*, IEEE, pp. 1–12.
- Ewing, B.T., Kruse, J.B., and Schroeder, J.L. (2006), Time series analysis of wind speed with time-varying turbulence, *Environmetrics*, 17(2), pp. 119–127.
- Glosten, L., Jagannathan, R., and Runkle, D. (1993), On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *The Journal of Finance*, 48, pp. 1779–1801.
- Haslett, J. and Raftery, A.E. (1989), Space-time modelling with long-memory dependence: Assessing Ireland’s wind power resource, *Applied Statistics*, 30(1), pp. 1–50.
- Hering, A.S. and Genton, M.G. (2010), Powering Up With Space-Time Wind Forecasting, *Journal of the American Statistical Association*, 105(489), pp. 92–104.
- Koopman, S.J., Ooms, M., and Carnero, M.A. (2007), Periodic seasonal reg-arfima-garch models for daily electricity spot prices, *Journal of the American Statistical Association*, 102(477), pp. 16–27.
- Le Guyader, C., Apprato, D., and Gout, C. (2014), Spline Approximation of Gradient Fields: Applications to Wind Velocity Fields, *Mathematics and Computers in Simulation*, 97, pp. 260–279.
- Ren, Y. and Zhang, X. (2010), Subset selection for vector autoregressive processes via adaptive Lasso, *Statistics & probability letters*, 80(23), pp. 1705–1712.
- Šaltyte Benth, J. and Šaltyte, L. (2011), Spatial-temporal model for wind speed in Lithuania, *Journal of Applied Statistics*, 38(6), pp. 1151–1168.
- Santos-Alamillos, F., Pozo-Vázquez, D., Ruiz-Arias, J.A. and Lara-Fanego, V., and Tovar-Pescador, J. (2014), A methodology for evaluating the spatial variability of wind energy resources: Application to assess the potential contribution of wind energy to baseload power, *Renewable Energy*, 69, pp. 147–156.
- Soman, S.S., Zareipour, H., Malik, O., and Mandal, P. (2010), A review of wind power and wind speed forecasting methods with different time horizons, *North American Power Symposium (NAPS), 2010*, IEEE, pp. 1–8.
- Taylor, J.W., McSharry, P.E., and Buizza, R. (2009), Wind power density forecasting using ensemble predictions and time series models, *Energy Conversion, IEEE Transactions on*, 24(3), pp. 775–782.
- Thapar, V., Agnihotri, G., and Sethi, V. (2011), Critical Analysis of Methods for Mathematical Modelling of Wind Turbines, *Renewable Energy*, 36, pp. 3166–3177.

- Tibshirani, R. (1996), Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.
- World Wind Energy Association, W.W.E.A. (2012), World wind energy report 2012, *WWEA: Bonn*.
- Wu, Y.K. and Hong, J.S. (2007), A literature review of wind forecasting technology in the world, *Power Tech, 2007 IEEE Lausanne*, IEEE, pp. 504–509.
- Zhu, X., Genton, M.G., Gu, Y., and Xie, L. (2014), Space-time wind speed forecasting for improved power system dispatch, *Test*, 23(1), pp. 1–25.
- Ziel, F., Steinert, R., and Husmann, S. (2015), Efficient modeling and forecasting of electricity spot prices, *Energy Economics*, 47, pp. 98–111.